DOCUMENT RESUME

ED 411 315 TM 027 381

AUTHOR Poggio, John P.; Glasnapp, Douglas R.; Green, Sam B.; Tollefson, Nona

TITLE Conducting Licensure Validity Studies: The Need To Broaden the Evidentiary Base.

PUB DATE 1997-03-00

NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS College Faculty; *Construct Validity; *Content Validity; Higher Education; *Licensing Examinations (Professions); *Teacher Certification; *Test Results; *Undergraduate Students

IDENTIFIERS ACT Assessment; Collegiate Assessment of Academic Proficiency; *Empirical Research; Pre Professional Skills Tests

ABSTRACT

Findings are reported from a series of comprehensive and detailed empirical investigations that evaluated four different teacher certification paper-and-pencil tests through a typical content validation method and from three distinct and independent empirical validation studies of these devices. These latter studies can be understood as providing construct-related evidence of validity. The four test batteries studied were: (1) the enhanced ACT Assessment; (2) the Collegiate Assessment of Academic Skills; (3) the Pre-Professional Skills Test; and (4) the state's locally developed comprehensive basic skills instruments. Three groups of participants provided data for the different types of studies conducted: 1,190 undergraduates, higher education faculty and administrators, and educators for kindergarten through grade 12. Content-related validity reviews were conducted by higher education faculty, and empirical studies involved test results from undergraduate students. Using both types of studies provided the opportunity of evaluating the appropriateness of sole reliance on the single, almost commonplace, content validation approach. These findings indicate that reliance on the content validation studies would have resulted in different recommendations than those that emerged from the broad-based empirical studies. (Contains four tables and four references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
John Poggio

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 411 315

Conducting Licensure Validity Studies: The Need to Broaden the Evidentiary Base

John P. Poggio
Douglas R. Glasnapp
Sam B. Green
Nona Tollefson

This paper is prepared for the:
Annual Meeting of the American Educational Research Association in Chicago, IL
March 1997

TM027381

Conducting Licensure Validity Studies: The Need to Broaden the Evidentiary Base*

John P. Poggio, Douglas R. Glasnapp, Sam B. Green and Nona Tollefson
University of Kansas

Abstract

In this paper we report on findings from a series of comprehensive and detailed empirical investigations that evaluated four different potential teacher certification paper and pencil tests through the lens of a typical content validation methodology as well as three distinct and independent empirical validation studies of these devices. The scope and nature of these latter studies can be understood as providing construct-related evidence of validity. Through this investigation we have had the opportunity to evaluate the appropriateness of sole reliance on the single, almost commonplace content validation approach. Findings reveal that sole dependence on this one approach appears ill advised rendering decisions not at all consistent with findings from the empirical investigations. Findings from the investigation are presented extensively in the paper and recommendations for such studies are discussed.

Introduction

School reform initiatives of the 1980s driven by dissatisfaction with the abilities and skills of persons entering teacher preparation and, in turn, with the skill levels of students graduating from the nations' high schools focused on entry level tests for teachers and exit level tests for students. Testing programs were viewed as the means by which better prepared teachers would enter classrooms and better prepared students would enter the work force. The ability to earn scores above a critical score was seen as the way to accomplish a set of diverse objectives. On one hand, tests were to serve a gate keeping position; people who did not possess certain requisite basic skills would not be admitted to teacher preparation programs. The testing program would, in turn, assure the general public that only truly qualified persons were being admitted to teacher preparation. Finally the introduction of entry level tests and end-of-preparation assessments such as licensure examinations would drive educational reform.

In no area has this preeminent role for assessment become more evident than in the reliance on tests for licensing, certification and employment decisions. Well over 75 percent of the states rely on a paper and pencil test for admission to initial certification programs or as an exit requirement in order to be recommended for a teaching license. On an altogether different level, California requires that persons already in the profession who wish to advance or maintain their certification pass a paper and pencil basic skills test. Testing has become the insurance policy for educational effectiveness as viewed by policy makers.

This era that has placed testing programs front and center in state educational reform efforts places a great responsibility on the educational measurement community. Evidence must be gathered that demonstrates that the instrument(s) used to make these individual high stakes decisions which are presumed, in turn, to drive school reform are valid. At the present time, considerable reliance is being placed on paper and pencil test's in admission and licensure decisions. However, the evidence required to support the appropriateness of

---

* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, 1997.

these decisions both in the short run, (i.e., ability to complete a program) and in the long run (i.e., judgments of effectiveness as a teacher) has focused solely on a single type of evidence, content-related validity evidence.

Even though heavy reliance has been placed on paper and pencil devices to yield these profound and everlasting decisions about an individual, the evidence required to support the appropriateness and trustworthiness of the measure used has focused solely on a single type, content-related validity evidence. Designs for establishing the content-related evidence of validity typically set out to confirm the properties of the measure; rarely is the content validation plan premised on disconfirming the otherwise apparent viability of the measure. Also, legal standards can be interpreted to only require minimal content related evidence; these now very dated standards (Uniform Guidelines, 1978 & 1979) are seen to suggest that criterion- or construct-related evidence is unnecessary to assert validity for an achievement measure. Only recently has the profession begun to question the suitability of reliance on the sole source content validity line of evidence (Camara & Brown, 1995; Madaus & Pullin, 1987; Messick, 1989; Poggio, Glasnapp, Miller, Tollefson & Burry, 1986).

The objective of the present inquiry was to evaluate the utility, suitability and need to extend validation investigations for teacher licensure tests beyond content related validity evidence. Findings are reported from a series of comprehensive and detailed empirical investigations that evaluated four different potential teacher certification paper and pencil tests through the lens of a typical content validation methodology as well as three distinct and independent empirical validation studies of these devices. Through this investigation the opportunity was provided to evaluate the appropriateness of reliance on the single content validation approach alone.

Methods and Procedures

Resulting from a joint effort of three state Education governing bodies, investigations were designed to evaluate the use of paper and pencil test scores in the basic skill areas of reading, writing and mathematics as criteria for admission to initial teacher education certification programs and to further assert the readiness of individuals to enter the profession. A series of validity studies examined and evaluated the appropriateness and utility of four different test batteries as initial screening measures for entrance to the teaching profession. The four test batteries studied included the: (1) Enhanced ACT Assessment tests marketed by American College Testing (ACT); (2) Collegiate Assessment of Academic Proficiency (CAAP) marketed by American College Testing (ACT); (3) Pre-Professional Skills Tests (PPST/Praxis I) marketed by Educational Testing Service (ETS); and, (4) the state's locally developed comprehensive basic skills assessment instruments (SKAT).

The purpose of an admission test for teacher certification programs is to serve as an initial screening measure on "basic skills" to help assure the applicant's readiness to adequately (successfully) fulfill professional requirements during their teacher preparation program and on the job as a teacher. Consequently, the series of validity studies were designed to assemble evidence that supported or refuted the appropriateness of each of the four tests to make inferences about the degree to which prospective teachers have the "basic skills" needed to be successful in their preparation program and on the job.

Three groups of participants provided data for the different types of studies conducted: a) currently enrolled higher education undergraduate students (1190); b) higher education faculty and administrators; and, c) K-12 education-related persons (teachers, administrators and Board of Education members). Coordination of the data collection was accomplished through mailed communications with contact persons at each of the 21 higher

education institutions in the state, with local school district superintendents and with individuals agreeing to participate. Three of the test batteries had released versions of the test, thus allowing for distribution of materials through the mail and unsupervised data collection at local sites. The available CAAP tests were secured forms and required that panels of raters be convened at central locations for supervised data collection.

Procedurally, data collection from higher education participants was initiated through contact with the Dean of Education at each of the 21 institutions with teacher certification programs. A contact person at each institution served as the coordinator of communication and data collection activities. The contact person provided data on number of education faculty and number of graduates in teacher education programs for the last two years. This information was used to proportionally sample faculty from each institution. An orientation meeting for the contact persons was held and all materials for data collection were distributed at this session or mailed to an institution's contact person for distribution.

For the K-12 education-related participants, initial contact was made with local district superintendents who were asked to nominate a predetermined number of teachers, administrators or the president or president-elect of the local Board of Education. A sample of those individuals nominated were sent packets of materials and a) were asked to review the PPST, ACT or SKAT tests or b) were invited to participate in one of the panel sessions held to review the CAAP tests. All K-12 participants were paid a stipend for conducting their reviews.

All review materials for the PPST, ACT and SKAT used in the content-related validity studies were self-directed and distributed by an institution's contact person or mailed directly to participants for completion individually at the local site. For the review and evaluation of the CAAP tests, a sample of higher education faculty and K-12 participants attended one of four review panel sessions held across the state.

Content-Related Validity Studies

Participants conducting the content-related validity reviews included higher education faculty with appointments in Schools of Education or teaching courses in which the majority enrollment would be students majoring in Education and K-12 education related persons.

Each participant received a packet of materials containing the review directions, response sheets and all subtests (Reading, Mathematics, Writing) for one of the four test batteries. For each basic skill content area subtest, two sets of ratings were requested, one set focusing on the individual test items and another set focusing on the skills measured by the test. In each instance, participants were asked to rate the extent to which the knowledge or skill measured by the specific item (or skill area) represents **essential prerequisite content knowledge or skill** for:

- performing at an adequate level in your teacher education curriculum, regardless of the area of specialization **(higher education referent).**
- performing adequately as a teacher in your school system regardless of the area of teaching specialization **(K-12 referent).**

The response scale used was 1) Not Necessary, 2) Limited Importance, 3) Important, and 4) Essential.

Table 1 lists the content areas and skills measured by each of the four tests reviewed. Eight sets of materials were developed, one for each of the four test batteries and different rating direction referents used for higher education faculty and K-12 participants.

Empirical/Construct Validity Studies

Two different data collection activities were implemented to more broadly address empirical oriented validity issues. One effort focused on actual student testing using reduced versions of the tests under review. Classes of undergraduate Education students completed the tests on a voluntary and cooperative basis at Higher Education institutions across the state. Copies of portions of the subtests from the three released test batteries were randomly sequenced and administered to students enrolled in classes instructed by Education faculty from cooperating institutions. The CAAP subtests were administered under secure conditions at one institution, but problems in data collection standardization and the fact that ACT would not release the scoring key makes the CAAP data collected from students unusable. Therefore, no information is provided on the CAAP tests as part of these empirical validation efforts.

In addition to having students take the tests, self-reported prior achievement information was obtained from each student, and current class achievement ratings were obtained from the students' instructors in the classes tested. For students judged as lower achieving students in the class, their instructors were asked to also judge the extent to which these students possessed the prerequisite knowledge and skills in reading, writing and mathematics needed to learn/master the information presented in class.

The second aspect of data collection for the construct focused validity studies obtained information from a college's student data base. Detailed instructions were sent to each institution contact person requesting that information for one-third of the students tested be obtained from the students' permanent records. The specific information sought data on the student's prior course enrollment history in the areas of reading/English, writing and mathematics and their GPA history in these courses and overall.

Results

Content-Related Validity Studies

Faculty in Institutions of Higher Education and K-12 education representatives participated in a content validation study of the four instruments: CAAP, PPST, ACT, and SKAT. The number of participants in each group are presented in Table 2. In total, data were provided by 151 higher education faculty and 81 K-12 persons. The content validation called for respondents to judge the relevance of: a) individual test items in Reading, Mathematics, and Writing, and b) skills assessed in each of the tested areas. Higher education faculty rated "the extent to which the knowledge or skill measured by the specific item represented essential prerequisite content knowledge or skill for performing at an adequate level in your teacher education curriculum, regardless of the area of specialization." K-12 education participants rated "the extent to which the knowledge or skill measured by the specific item represented essential prerequisite content knowledge or skill for performing adequately as a teacher in your school system regardless of the area of teaching specialization." The ratings were make using a 4-point scale of "Not Necessary," "Limited Importance", "Important", and "Essential."

Mean item ratings were computed for the faculty and K-12 groups and the percentage of respondents in each group who rated each item "Important" or "Essential" was determined. In any individual validity study, sampling and measurement errors are likely,

and, for this reason, the critical value for assessing the content validity of an item was set at 64 percent for higher education faculty and 69 percent for the K-12 persons in the present study. This critical value was computed as 50 percent plus 1.65 times the sampling standard error of the proportion. Establishing such critical values increases the confidence one can have that at least 50 percent of a different respondent group would judge the content measured in a similar number of items from each test to be "Important" or "Essential".

Table 2 reports the number and percentage of items receiving different levels of endorsement by faculty and K-12 participants. Frequency distributions of items based on the level of endorsement are presented by the content area measured. Using the CAAP Reading test data as an illustration, all 36 items (100 percent) met the criteria of being endorsed by at least 65 percent of the faculty group and 31 items (86 percent) met the criteria of being endorsed by at least 70 percent of the K-12 education group.

Inspection of the frequency distributions for the Reading tests shows that both the faculty and K-12 education groups judged the Reading tests of the CAAP, the PPST, and the SKAT to be content valid. For all of these tests, more than 80 percent of the items were judged "important" or "essential" by at least 70 percent of the respondent groups. For the ACT, the K-12 education sample judged the items to be more valid for "performing adequately as a teacher" than the faculty sample judged the items to be necessary for successfully completing a teacher preparation program.

The writing tests were judged content valid by both the faculty and K-12 education samples. For all tests, at least 80 percent of the items met the criteria of being endorsed by at least 70 percent of the faculty and K-12 education respondents.

The mathematics tests were judged to be less content valid than either the reading or writing tests. The PPST was the only instrument for which at least 75 percent of the items met the criteria for endorsement. Items on the ACT mathematics test had the lowest content validity ratings. About half of the items had endorsement rates less than the criterion for the K-12 education sample and about three-fourths of the items had endorsement rates less than the criterion for the faculty sample.

For the four tests studied, the PPST had the highest overall percentage of items meeting the content-related validity endorsement criteria. At least 75 percent of the items on all PPST subtests met the criteria of being endorsed by at least 70 percent of the respondents. The SKAT had the second highest item validity ratings overall. The skills on at least 80 percent of the items on the Reading and Writing Tests were judged important or essential by at least 70 percent of the faculty and K-12 education groups. The SKAT mathematics items were rated lower than either the reading or writing items. Between 40 percent and 60 percent of the SKAT mathematics items achieved the critical value of at least 70 percent of the respondents judging the item as important or essential. The CAAP items were rated slightly lower than the SKAT items, while the ACT items received the lowest content validity ratings overall. The highest validity ratings were assigned to items on the ACT Writing test and the lowest validity rating were assigned to items on the ACT Mathematics test.

Table 3 reports the frequency and percentage of the faculty and K-12 education samples who judged the skills measured by each of the subtests as either "important" or "essential". Inspection of the data in Table 3 shows that skills measured by all of the Reading Tests were judged important or essential by at least 70 percent of the faculty and K-12 education groups. The mathematical skills measured by the CAAP and SKAT also were judged important or essential by at least 70 percent of the raters. For the ACT, basic algebra skills were judged important and/or

essential, but advanced algebra, trigonometry, and calculus skills were considered not necessary or of limited importance. Writing skills assessed by all four of the tests, like reading skills, were judged important or essential by at least 70 percent of the faculty and K-12 education groups.

Empirical Construct-related Validity Studies

Data for the criterion related studies include actual student performance on the SKAT, PPST and ACT Reading and Mathematics multiple-choice tests and the PPST and ACT Writing multiple-choice tests as the primary scores. The CAAP tests were not included in these studies.

Data from students currently enrolled in a school of education or students who expressed an intention to enroll in a school of education was the only data used in the criterion-related validity studies. Table 4 shows the number of students completing the reduced forms of each of the tests. In all cases at least 100 students completed each reduced form of the test.

Six criterion measures were used in these studies. Two measures represented prior educational attainments; they were GPA in high school mathematics courses and GPA in high school English courses. Four were college level achievement measures including: students' self-reported GPA in education courses, highest level mathematics course taken, college instructors' ratings of course content achievement, and cumulative GPA. The first five criterion measures were obtained from student self-reports. The sixth measure, cumulative GPA, was obtained from students' records Sample sizes for the transcript analysis data were, on average, 36 students for each reduced form of the test.

Table 4
Number of Students Completing Each of the Reduced Forms of the SKAT. ACT, and PPST

Test Form	Number of Students
SKAT Reading	105
SKAT Mathematics	137
ACT Reading	117
ACT Writing	130
ACT Mathematics	134
PPST Reading	125
PPST Writing	136
PPST Mathematics	137

The relationship between student performance on the tests and the achievement criterion measures are presented in Table 5 as Pearson correlation coefficients. Coefficients above .35 are considered acceptable indicators of adequate criterion-related validity in studies such as the present one. Coefficients in the .20 to .34 range are statistically significant given the present sample sizes, but are considered as presenting borderline evidence as indicators of validity.

The evidence in Table 5 lends credence to the content-related validity results for the PPST, ACT and SKAT Reading and Writing multiple-choice tests. The relationships between actual student performance on the tests and other indicators of achievement (i. e., instructor ratings, grade point average in education courses, and high school English grade

point average) were in the expected direction and followed a consistent pattern across tests. The data for the PPST Reading test and the ACT Writing test were particularly strong with the remaining coefficients supportive, but in a borderline range as validity indices.

The data on the Mathematics tests are consistent with the content review results for the ACT and SKAT tests in that a lack of supporting evidence is observed. Little or no relationship was found between actual student performance and the other indicators of achievement in education program courses (i. e., instructor ratings or grade point average in education courses). This lack of relationship also was observed for the PPST Mathematics test. While this latter test was the only Mathematics test to receive support from the content-related validity studies, the criterion-related validity indices offer limited support for its use.

Examining the relationship of student performance on the tests under study with data from students' files/transcripts was intended to provide additional direct evidence from the enrollment and achievement history of students that would contribute to the validity arguments for or against the use of a particular test. Additionally, these data and patterns of relationships help verify the trustworthiness of the self-reported and test performance data from students. In general, the magnitude and pattern of relationships confirm the trustworthiness of the data in that they are of an acceptable magnitude and in the directions expected. For example, the correlation between student self-reported grades in education classes and their transcript cumulative grade point average was .685. Correlations between student self-reported high school GPAs in English and Math courses with their transcript college cumulative GPA were .357 and .351, respectively.

Transcript data on the college enrollment pattern and performance in required English and Mathematics courses proved not to be convincing in either direction when addressing the validity of the individual tests. Surprisingly for the approximately 330 students on which transcript data were secured, an English course requirement prior to admission to a School of Education was reported for 86 percent of the students. A mathematics course admissions requirement was reported for only 24 percent of the students. This is not to infer that satisfactory completion of a mathematics course is not a degree or certification requirement for students in education. The mathematics course completion just is not an admissions requirement for many programs. For this reason, the transcript grade performance data on the grade in highest level college mathematics course was too limited to provide stable estimates of relationships to performance on the individual tests.

When collapsed across test samples, however, the correlations of grade in highest level English and highest level mathematics courses taken with instructors' ratings of student in-class performance level were .257 (n=260) and .237 (n=72), respectively. For the grade in highest level English course, the correlations with actual student performance on the different reading and writing tests were: .114 for the SKAT reading test (n=30), .400 for the ACT reading test (n=32), .320 for the ACT writing test (n=36), .316 for the PPST reading test (n=29), and .133 for the PPST writing test (n=38).

Other relevant file/transcript data to decision making is the relationship between the students' cumulative college GPA and instructors' class performance ratings. This relationship was .584 (n=285), thus providing the necessary confidence that the instructor ratings are reasonable and trustworthy measures of students' levels of performance across a variety of classes taken by education students.

While such construct-related validity evidence for the Mathematics tests offers only weak support for their use as admissions tests to teacher education programs, the correlations offer evidence that these Mathematics tests are valid measures of mathematical

skills. Sufficiently high correlations were observed between test performance on each of the mathematics tests and either high school mathematics grade point average or highest level mathematics course taken to document a relationship between test performance and achievement in mathematics. The skills measured were not considered sufficiently important or essential to "performing at an adequate level in the teacher education curriculum," or "performing adequately as a teacher in a school system." An alternative to requiring students to pass a mathematics skills would be to require that students satisfy a course curriculum standard. That is, require students to complete a higher level mathematics course at either the high school or college level.

Education students' performance on the reduced forms of the three tests was at reasonably high levels. The average performance of education students sampled was above the norm group averages for all tests studied. For the state SKAT tests, the mean performance on the Reading test was 70 percent of the items correct compared to a grade 10 normative mean performance of 69 percent correct. In mathematics, the mean performance was 59 percent correct compared to a grade 10 normative mean performance of 40 percent correct. Best estimates for comparative performance on the ACT and PPST indicate that the average student performance rates on the ACT would translate into ACT scale scores of 21 in each of the three content areas tested and for the PPST would translate into scale scores of 176 in mathematics and 178 in reading.

As an additional supplementary piece of information, when instructors in education courses were asked to rate the achievement level of students in class, 78 percent were judged to be achieving course content at a high level (B- or higher). For the low achieving students (C+ or lower; n=260), only 23 students were judged unequivocally not to have the necessary prerequisite knowledge and skills in reading, writing and mathematics needed to learn/master the information presented in class. These data would indicate that instructors certainly feel that the vast majority of current students have the necessary prerequisite knowledge and skills in reading, writing and mathematics needed to learn/master the information presented in class.

Summary and Conclusions forom the Investigation

Based on analyses of data gathered in the conduct of the content-related validity and construct-related validity studies, the following results were observed to guide decisions.

1. Support for the content validity of all four reading and writing tests was evident. All skills measured by the tests were judged important or essential by all reviewers of the examinations. At the test item level, K-12 educators gave strong support for the skills measured by the test items on all four tests, while higher education faculty gave strong support for the items on the PPST, CAAP, and the state's tests and lesser, but still acceptable, support for ACT reading test questions.

2. Content validity evidence only supported the mathematics test of the PPST. The general mathematical skills (e.g. pre-algebra, algebra, etc.) measured by the PPST, CAAP and state tests were judged to be important or essential. The ACT general advanced skills (advanced algebra, trigonometry, and calculus) were considered not necessary or of limited importance to the purposes of performance in Education classes and growth upon entry to the profession. When reviewed at the item level, at least 75 percent of the items on the PPST were endorsed by the two rating groups for these purposes. For the ACT, only 30 percent of the items were endorsed. For the CAAP and state tests, only 37 and 44 percent of the items were endorsed, respectively. In summary content validity data supported only the PPST mathematics subtest.

3. The data gathered from the empirical validity studies confirmed the content-related validity results for the PPST, ACT and state's reading and writing multiple-choice tests. The relationships between actual student performance on the tests and other indicators of achievement (i. e., instructor ratings, grade point average in Education courses, high school English grade point average, and cumulative and semester grade point averages in college) were in the expected direction, attained reasonable levels of magnitude and followed a consistent pattern across tests. The data for the PPST reading test and the ACT writing test were particularly strong with the remaining coefficients supportive, but in a borderline range as validity indices.

4. The data on the mathematics tests available from the criterion-related validity studies are consistent with the content review results for the ACT and state tests in that they did not support the use of these skill area tests. A weak relationship was found between actual student performance on these tests and the other indicators of achievement in Education program courses. A weak relationship also was observed for the PPST mathematics test. While the PPST test was the only mathematics test to receive support from the content-related validity studies, the criterion-related validity indices does not offer convincing support for its use.

5. While the content validity evidence for the ACT, CAAP and state's mathematics tests are not supportive of their use as admissions tests to initial teacher education certification programs and the teaching profession, evidence became available that documents these mathematics tests are valid measures of mathematical skills. This is neither a contradiction or an inconsistency. The knowledge and skills measured as reflected in the questions on these tests are not considered sufficiently important or essential to "performing at an adequate level in the Teacher Education curriculum," or "performing adequately as a teacher in a school system." However, persons possessing advanced mathematics skills do demonstrate higher levels of achievement in college course work as reflected by GPA indicators.

   While the content validity evidence does not suggest that scores on the mathematics portions of the tests investigated should be used to make admission decisions, the relationships between scores in mathematics and performance in high school and college were strong. Student performance on the mathematics tests, were related to high school GPA, highest level math course taken, college math classes taken, grades received, and cumulative GPAs. For this reason, it is recommended that students show competence in mathematics by meeting a defined, explicit course curriculum standard that is defined as passing a higher level mathematics course at either the high school or college level. This recommendation is offered based on the investigations involving file transcript information and correlates to test performance.

6. Beginning from the premise that all teachers need to demonstrate basic competence in mathematics, equating studies could be undertaken that would establish cut scores on the ACT, and perhaps the CAAP, that corresponded to the cut score on the PPST. As many colleges and universities in the state either require or encourage entering freshmen to complete the ACT, equating ACT scores to PPST scores would be both efficient and cost effective for both students who want to enter teacher preparation programs and the institutions they attend..

7. The data on currently enrolled Education program students from the criterion-related validity studies supply unique information that needs to affect decision making. First, actual student performance for those taking the tests was at a reasonably high level for all tests, (i. e., the average performance of Education students sampled was well above

the norm group averages). Second, when instructors in Education courses were asked to rate the achievement level of students in class, 78 percent were judged to be achieving course content at a high level in their course (grades of B- or higher). For the lower achieving students (n=260), only 23 (2%) students were judged unequivocally not to have the necessary prerequisite knowledge and skills in reading, writing and mathematics needed to learn/master the information presented in class.

Educational Significance

Broad based validity studies provide a rich set of data for policy makers to use in making decisions. The findings of this study demonstrate that traditional content validation strategies would have lead to different recommendations than those that emerged from the more comprehensive content and empirical validation strategies that were conducted. Furthermore, such broad-based studies produced unexpected outcomes that pointed to the need for additional data collection and study. When tests are used to make decisions that have long term consequences for students' lives and their futures, policy makers need to have information that helps them to understand the types of inferences that can reasonably be made from individual's scores and from groups of individuals' scores. It is our opinion that standards for demonstrating the validity of achievement tests whose scores are used to make high stakes admission and licensure decisions should include evidence of both content-related and empirical validity before these types of test interpretations are made.

References

Camara, W. J., & Brown, D. C. (1995). Educational and employment testing: Changing concepts in measurement and policy. Educational Measurement: Issues and Practice, 14 , 5-11.

Madaus, G.F. and Pullin, D. (1987). Teacher Certification Tests: Do they really measure what we need to know. Phi Delta Kappan, September, 31-44.

Messick, S. (1989). Validity. In R.L.Linn (Ed.), Educational Measurement (3rd ed., pp. 13-103). New York: Macmillan.

Poggio, J.P., Glasnapp, D.R., Miller, M.D., Tollefson, N. and Burry, J.A. (1986). Strategies for validating teacher certification tests. Educational Measurement: Issues and Practice, 18-25.

Table 1
List of Content Area Skills Measured by the Tests

	ACT	CAAP	PPST	SKAT
Reading	Referring Reasoning	Referring to what is explicitly stated Reasoning to determine implicit meaning Drawing conclusions, comparisons, and generalization beyond the text	Main Ideas/ Primary Purposes Supporting Ideas Organization Vocabulary in Context Evaluating Arguments Inferential Reasoning Generalization	Narrative Reading Comprehension Expository Reading Comprehension
Mathematics	Pre-Algebra Elementary Algebra Intermediate Algebra Coordinate Geometry Plane Geometry Trigonometry	Pre-Algebra & Elementary Algebra Skills Intermediate Algebra and Coordinate Geometry Advance Algebra Trigonometry Calculus	Conceptual Knowledge Procedural Knowledge Comprehension of Representations of Quantitative Information Formal Mathematical Reasoning Ability	Knowledge Base Non-Routine Problem Solving Mathematical Reasoning Communication Mathematical Conceptualization Mathematical Procedures Implementation (content areas of number operations, measurement, geometry, graph and data, probability and statistics, functions, patterns, logic)
Writing	Punctuation Basic Grammar & Usage Sentence Structure Rhetorical Strategy Organization Style	Punctuation Grammar Sentence Structure Organization Strategy Style	Structure and Grammar Idioms and Word Choice Mechanics No Error	No test
Writing Essay	No test	Formulating an Assertion Supporting the Assertion Organizing and Connecting Major Ideas Clearly Expressing Ideas	Appropriate for the Assigned Task and for the Intended Audience Organize and Develop Ideas Logically Clear Focus Supporting Reasons, Examples, Details, Etc. Sentence Variety and Facility Free of Errors in Standard Written English	Ideas and Content Organization Voice Word Choice Sentence Fluency Writing Conventions

BEST COPY AVAILABLE

Table 2
Number and Percentage of Items Receiving Different Levels of Content-Related Validity Endorsement

	% of Raters Endorsing Item	CAAP		PPST		ACT		SKAT	
		High. Ed. n = 42	K-12 n = 20	High. Ed. n = 38	K-12 n = 22	High. Ed. n = 34	K-12 n = 15	High. Ed. n = 37	K-12 n = 24
Reading	> 70	35 (97)*	31 (86)	35 (88)	36 (90)	25 (63)	34 (85)	21 (95)	18 (82)
	64 - 69.99	1 (3)	3 (8)	5 (12)	2 (5)	3 (7)	4 (10)	1 (5)	2 (9)
	60 - 63.99	0 (0)	2 (6)	0 (0)	1 (3)	2 (5)	1 (3)	0 (0)	1 (5)
	50 - 59.99	0 (0)	0 (0)	0 (0)	0 (0)	7 (18)	1 (2)	0 (0)	1 (4)
	< 50	0 (0)	0 (0)	0 (0)	0 (0)	3 (7)	0 (0)	0 (0)	0 (0)
Mathematics	> 70	11 (31)	12 (34)	31 (78)	30 (75)	11 (18)	18 (30)	28 (58)	21 (44)
	64 - 69.99	2 (6)	2 (6)	1 (2)	4 (10)	3 (5)	5 (8)	4 (9)	7 (14)
	60 - 63.99	1 (3)	3 (9)	2 (6)	0 (0)	2 (3)	2 (3)	2 (4)	5 (10)
	50 - 59.99	7 (20)	4 (11)	1 (2)	3 (8)	0 (0)	6 (10)	5 (10)	7 (14)
	< 50	14 (40)	14 (40)	5 (12)	3 (8)	44 (74)	29 (49)	9 (19)	9 (18)
Writing	> 70	71 (99)	63 (87)	45 (100)	45 (100)	65 (87)	69 (92)	-- no test --	
	64 - 69.99	1 (1)	3 (4)	0 (0)	0 (0)	6 (7)	6 (8)		
	60 - 63.99	1 (2)	0 (0)	0 (0)	0 (0)	1 (2)	0 (0)		
	50 - 59.99	0 (0)	4 (6)	0 (0)	0 (0)	1 (2)	0 (0)		
	< 50	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)	0 (0)		
Writing Essay	> 70	2 (100)	2 (100)	3 (100)	3 (100)	-- no test --		6 (100)	5 (83)
	64 - 69.99	1 (1)	3 (4)	0 (0)	0 (0)			0 (0)	0 (0)
	60 - 63.99	1 (2)	0 (0)	0 (0)	0 (0)			0 (0)	0 (0)
	50 - 59.99	0 (0)	4 (6)	0 (0)	0 (0)			0 (0)	1 (17)
	< 50	0 (0)	0 (0)	0 (0)	0 (0)			0 (0)	0 (0)

* Percent of items

Table 3
Number and percentage of content area skills receiving content-related validity endorsement

	High. Educ.	K-12
Reading		
CAAP	3 (100)*	3 (100)
PPST	7 (100)	7 (100)
ACT	2 (100)	2 (100)
SKAT	2 (100)	2 (100)
Mathematics		
CAAP	2 (40)	2 (40)
PPST	5 (100)	5 (100)
ACT	3 (50)	2 (33)
SKAT	6 (100)	6 (100)
Writing		
CAAP	5 (83)	5 (83)
CAAP-Essay	2 (100)	2 (100)
PPST	5 (84)	6 (100)
ACT	6 (100)	6 (100)
SKAT	5 (83)	6 (100)

* Percent of skills